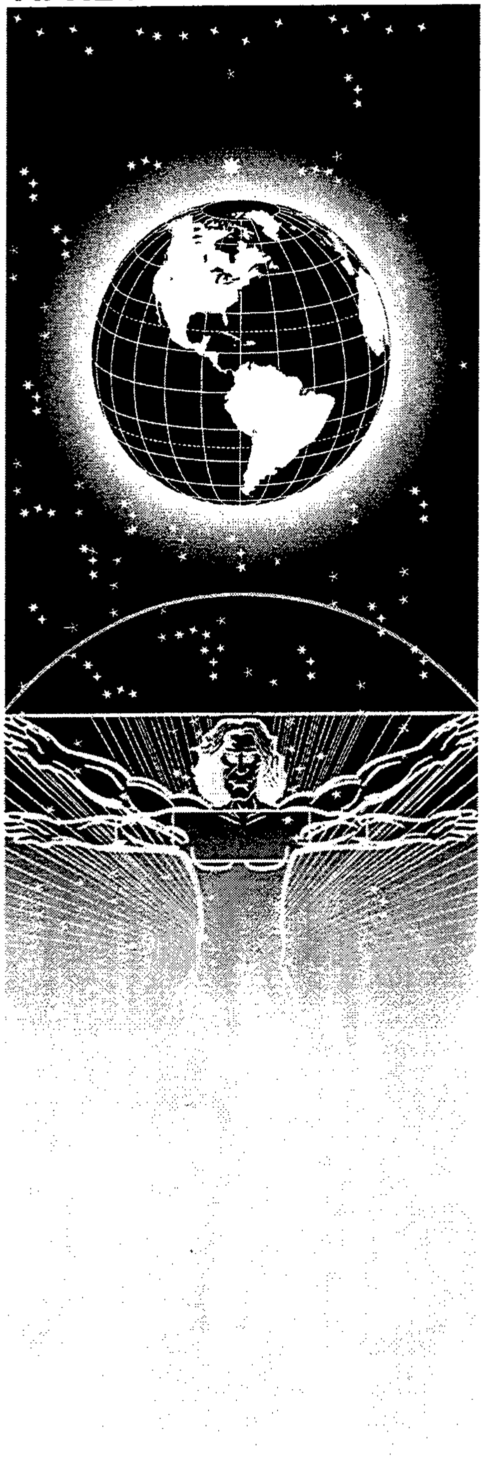
AFRL-HE-WP-TR-2002-0019



UNITED STATES AIR FORCE RESEARCH LABORATORY

APPLICATION OF ARTIFICIAL NEURAL NETWORKS FOR AIR TRAFFIC CONTROLLER FUNCTIONAL STATE CLASSIFICATION

Chris A. Russell
Glenn F. Wilson

HUMAN EFFECTIVENESS DIRECTORATE
CREW SYSTEM INTERFACE DIVISION
WRIGHT-PATTERSON AFB, OHIO 45433-7022

JUNE 2001

INTERIM REPORT FOR THE PERIOD JANUARY 2001 TO JUNE 2001

Approved for public release; distribution is unlimited.

Human Effectiveness Directorate
Crew System Interface Division
2255 H Street
Wright-Patterson AFB OH 45433-7022

NOTICES

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Please do not request copies of this report from the Air Force Research Laboratory. Additional copies may be purchased from:

National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161

Federal Government agencies and their contractors registered with the Defense Technical Information Center should direct requests for copies of this report to:

Defense Technical Information Center
8725 John J. Kingman Road, Suite 0944
Ft. Belvoir, Virginia 22060-6218

TECHNICAL REVIEW AND APPROVAL

AFRL-HE-WP-TR-2002-0019

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public.

The voluntary informed consent of the subjects used in this research was obtained as required by Air Force Instruction 40-402.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

WILLIAM C. SIMON, Lt Col, USAF, BSC
Deputy Chief, Crew System Interface Division
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 2001	Interim Report, January 2001 to June 2001

4. TITLE AND SUBTITLE

Application of Artificial Neural Networks for Air Traffic Controller Functional State Classification

5. FUNDING NUMBERS

PE 62202F
PR 7184
TA 08
WU 64

6. AUTHOR(S)

Chris A. Russell
Glenn F. Wilson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory
Human Effectiveness Directorate
Crew System Interface Division
Air Force Materiel Command
Wright-Patterson AFB, OH 45433-7022

8. PERFORMING ORGANIZATION REPORT NUMBER

AFRL-HE-WP-TR-2002-0019

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

Determining operator cognitive or functional state is a critical component of adaptive aiding systems. To determine cognitive state, we must decide which measured features will assist in distinguishing different levels of mental activity. Psychophysiological signals were collected for two levels of cognitive workload from which 43 measures were derived. Three feature reduction methods were applied, and the results were used as inputs to an artificial neural network for training and classification. Average classification accuracies up to 89.7% were achieved and the number of input features required was reduced by up to 84 percent.

14. SUBJECT TERMS	15. NUMBER OF PAGES
artificial neural networks, cognitive workload, feature saliency, psychophysiological measures, air traffic control, pattern classification	52
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

THIS PAGE INTENTIONALLY LEFT BLANK.

ABSTRACT

Cognitive workload for seven air traffic controllers was estimated using backpropagation feedforward artificial neural networks (ANN). Multiple channels of eye movement corrected, continuous electroencephalograph (EEG) recordings, eye blink activity, heart rate and respiration intervals were used as input features to classify four levels of mental workload in a simulated air traffic control study. The workload levels represented were low, medium and high as well as an overload condition. Workload levels were manipulated by changing the volume of aircraft or the complexity of the task. Salient psychophysiological features were determined using a partial derivative method providing an input-output relationship for each feature.

The data were evaluated as a seven-class, four-class or a two-class problem. The seven-class problem consisted of low, medium and high conditions for both the volume and complexity manipulations and the overload condition. The overall mean classification accuracy was 80 percent across seven controllers.

The four-class problem separated the manipulations of volume and complexity as two distinct data sets. Both data sets consisted of low, medium and high conditions plus the overload condition. A mean classification accuracy of 84% for seven controllers is reported. Feature reduction consisted of removing the non-salient features from the data set. Reducing the feature set from 88 input features to the nineteen most salient input features increased the mean classification accuracy from 84% to 93% for the four-class problem.

The two-class problem combined the low, medium and high volume data as one class of workload and the low, medium and high complexity data as one class of workload. Each were compared to overload class. An average of 98% classification accuracy across all seven subjects resulted from using the two-class problem. An average of eight features was used after feature reduction. Psychophysiological data used with ANNs can very accurately classify air traffic controller cognitive workload. Application of these procedures to cognitive workload evaluation and adaptive aiding shows tremendous promise.

TABLE OF CONTENTS

INTRODUCTION 1

METHODS 3

- Subjects 3
- Simulator Task 3
- Data Collection 5

PROCEDURE 6

- Feature Selection 6
- Artificial Neural Network 8
- Feature Reduction 14

RESULTS 18

- Seven-class Problem Using All Features 18
- Four-class Problem Using All Features 19
- Four-class Problem Using Reduced Features 22
- Two-class Problem Using Reduced Features 26

DISCUSSION 31

REFERENCES 36

APPENDIX A – Individual Subject Probability Matrices Using All Features 39

APPENDIX B – Individual Subject Probability Matrices Using Reduced Features 41

APPENDIX C – Seven-Class Individual Subject Probability Matrices 43

LIST OF FIGURES

Figure 1. Sample TRACON Display 4

Figure 2. Frequency Response of Elliptical Filters 6

Figure 3. Description of Moving Window 7

Figure 4. Network architecture showing a fully connected network with the number of neurons in each layer. The form of the logistic sigmoid activation function is provided at the bottom 8

Figure 5. Individual neuron showing the weighted sum of the inputs followed by the logistic sigmoid activation function, *f(a)*. 9

Figure 6. Classification behavior of the ANN as features are removed 17

Figure 7. Individual Classification Accuracy 19

Figure 8. Individual Subject Classification Accuracy. 21

Figure 9. Overload Condition Classification Accuracy. 22

Figure 10. Subject classification accuracy after reduction for the four-class problem. 23

Figure 11. a) Salient electrode sites for four-class problem. b) Salient electrode sites for volume four-class problem. c) Salient electrode sites for complexity four-class problem. 25

Figure 12. Overlap of electrode sites for four-class reduced feature analysis. 25

Figure 13. Individual subject classification accuracy for the two-class problem. 27

Figure 14. a) Salient electrode sites volume and complexity. b) Salient electrode site for volume data. c) Salient electrode site for complexity data. 29

Figure 15. Overlap of Electrode Sites. 30

LIST OF TABLES

Table 1. Seven-class Probability Matrix 19

Table 2. Complexity Data Probability Matrix 20

Table 3. Volume Data Probability Matrix 20

Table 4. Complexity Probability Matrix 23

Table 5. Volume Probability Matrix 23

Table 6. Salient Features – Volume 24

Table 7. Salient Features – Complexity 24

Table 8. Salient Features – Volume and Complexity 24

Table 9. Two-class Volume Accuracy 26

Table 10. Two-class Complexity Accuracy 26

Table 11. Salient Features – Volume and Complexity 28

Table 12. Salient Features – Volume 28

Table 13. Salient Features – Complexity 28

INTRODUCTION

Air traffic controllers have long considered their jobs to be one of the most cognitively challenging, demanding and stressful jobs in the world. In fact, the Federal Aviation Administration has been introducing automation and improved navigational equipment to reduce the workload of controllers and improve the management of airspace (Benel, Dancy, Dehn, Gutmann and Smith, 1989; Lee, Pawlak, Sanford and Slattery, 1995; Perry, 1997). Increased traffic in our airspace, as well as expanding airport capacities, has resulted in an increase in controller errors. These errors are manifested as monitoring failures, wrong heading and altitude assignments and improperly executed hand-offs (Morrison & Wright, 1989).

Developing measures of workload is key to evaluating workload savings through automation and equipment upgrades. Subjective measures, such as the NASA-TLX (Hart and Staveland, 1988) and SWAT (Reid & Nygren, 1988), are viable metrics. However, they require the operator to report how hard he or she is working. Typically, interrupting the task and instructing the operator to complete a questionnaire regarding mental and physical work accomplish this. This is invasive, sometimes prone to bias, and is difficult to use in real-time.

Psychophysiological measures such as EEG, heart rate, eye movement and blink rates, and respiratory rates are non-invasive and can be computed real-time. These measures have been used to discriminate cognitive workload in several tasks (Greene, Bauer, Kabrisky, Rogers, Russell, and Wilson, 1996; Russell, Monett and Wilson, 1996).

This study investigates the ability of an artificial neural network (ANN) to correctly classify several levels of cognitive workload, including cognitive overload, in an air traffic control simulation. An approach to classification using ANNs seems appropriate because

humans are complex, nonlinear systems. ANNs are nonlinear and do not make assumptions about the distributions of the data. Neural networks have been used in classification of cognitive workload in several studies. Anderson, Devulapalli, and Stolz (1995) investigated single task workload classification using alpha band activity and autoregressive methods. Gevins, Smith, Leong, McEvoy, Whitfield, Du and Rush (1998), used EEG with ANN classifiers, manipulated low, moderate and high working memory load states and compared each load pair in the classification process. Cognitive workload estimation was investigated using EEG activity with ANNs during a simulated aircraft landing task (Russell, Monett and Wilson, 1996, Greene, Bauer, Kabrisky, Rogers, Russell and Wilson, 1996), and in an air-to-ground Scud hunt mission (Russell, Reid and Vidulich, 2000).

Physiological data, in addition to being nonlinear, are neither normally distributed nor stationary and the true distribution is not known. Fishers' discriminant analysis treats diagnostic tests as multivariate normal and the covariance matrices of the test are assumed to be equal. ANNs have advantages in that they are distribution free, meaning that the statistical distribution of the data is not important. This means clear superiority over classical statistical methods when there is no knowledge of distribution function or if the data are non-Gaussian.

Much of the previous work in this area evaluated the use of ANNs in single task environments (Anderson, Devulapalli, & Stolz, 1995, Gevins, Smith, Leong, McEvoy, Whitfield, Du & Rush, 1998). Single task experimentation provides the foundation for implementing artificial neural networks in cognitive workload classification. It shows that ANNs can discriminate patterns in psychophysiological data. However, in the real world most of the problem areas associated with cognitive overload are found in multitask environments such as air traffic control and piloting aircraft.

METHODS

Subjects

Seven U.S. Air Force air traffic controllers were used in this study. Subjects ranged from 21 to 29 years of age and were all right-handed. Their experience levels in air traffic control ranged from 2.5 to 7.5 years (Brookings, Wilson and Swain, 1996).

Simulator Task

TRACON for Windows (Version 1.03), an air traffic control simulation created by Wesson International, was used in this study. The simulation display (see Figure 1) was comprised of four elements, a color radarscope of Los Angeles International airport and four surrounding airports, a communications display consisting of controller commands and pilot responses, flight strips representing active and pending aircraft, and the controller's score for the current scenario.

The workload levels were manipulated by increasing the volume of aircraft or the complexity of the situation presented to the subject. Three levels of workload were evaluated, low, medium and high. The volume condition consisted of manipulating the number of aircraft presented to the subject over the session. The number of aircraft was six for the low condition, twelve for the medium level, and eighteen for the high workload condition. The aircraft were presented in a fifteen-minute time interval for each of the conditions. The complexity conditions were simulated by varying the traffic mix presented to each subject while maintaining the number of aircraft constant at twelve. These manipulations were the result of varying the aircraft

types and the ratio of arrivals and departures. Finally, for the overload condition fifteen aircraft were presented to each controller during a five-minute period.

The NASA-TLX was used to collect subjective estimates of workload for each condition. Six subscales are collected from the NASA-TLX subjective workload score. They are mental demand, physical demand, temporal demand, performance, effort, and frustration. A composite TLX score is computed from a combination of the six subscales. The TLX results verified that four separate difficulty levels were achieved.

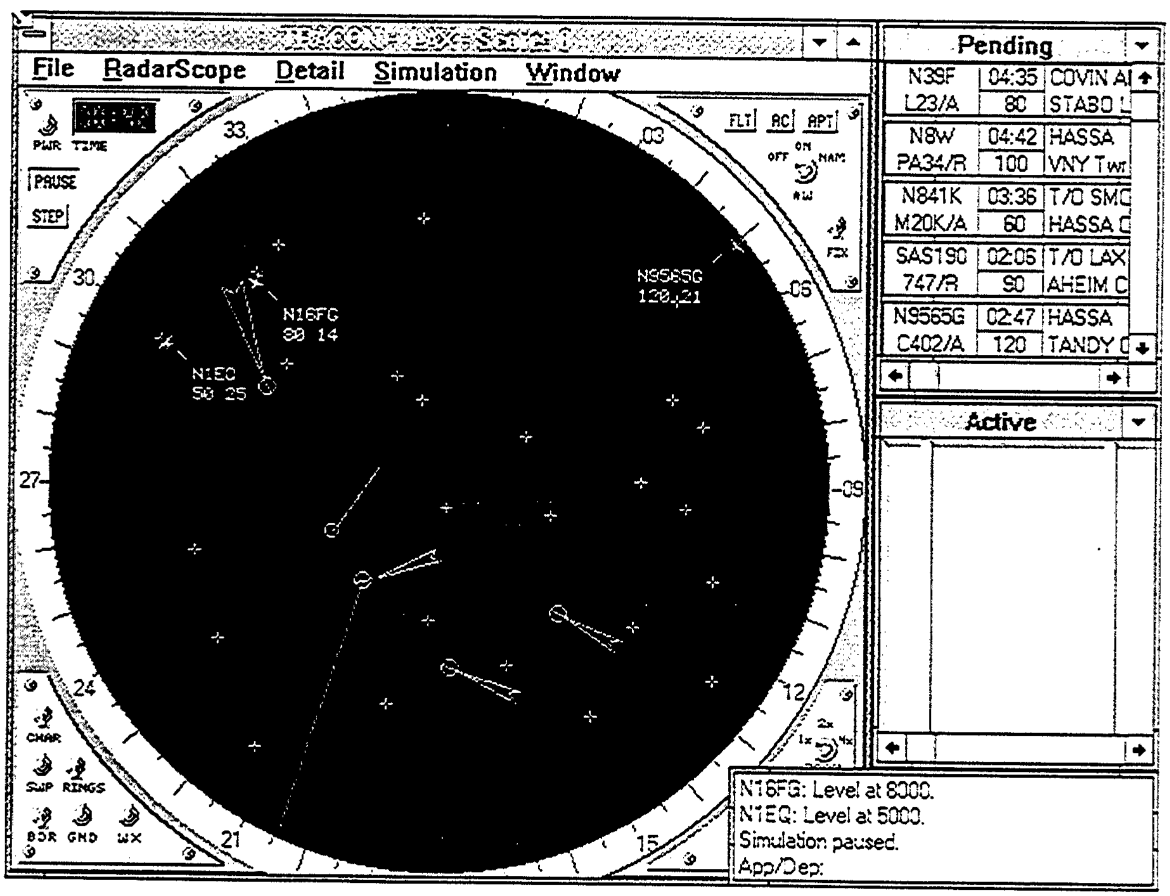


Figure 1. Sample TRACON Display

Data Collection

Nineteen channels of EEG data were recorded at sites positioned according to the International 10-20 electrode system (Jasper, 1958) using a Biologic Brain Atlas III and an ElectroCap. Mastoids were used as references. Electrode impedences were below 5K ohms. The amplifier gain was 30,000 and the data were passed through a bandpass filter with cutoff frequencies of 0.1 and 30 Hz. Eye blinks, heart rate, and respiration intervals were also collected.

PROCEDURE

Feature Selection

The data from each electrode site was filtered using a bank of elliptical filters to produce five bands of EEG: delta (DC-3 Hz), theta (4-7 Hz), alpha (8-12 Hz), beta (13-30 Hz) and gamma (31-42 Hz). Elliptical filters are ideal since they have sharp cutoff frequencies and low order. The filters used in this study were eighth order elliptical filters with stopband attenuation of 20 dB and a passband ripple of 1 dB. Figure 2 shows the frequency response of the elliptical filters.

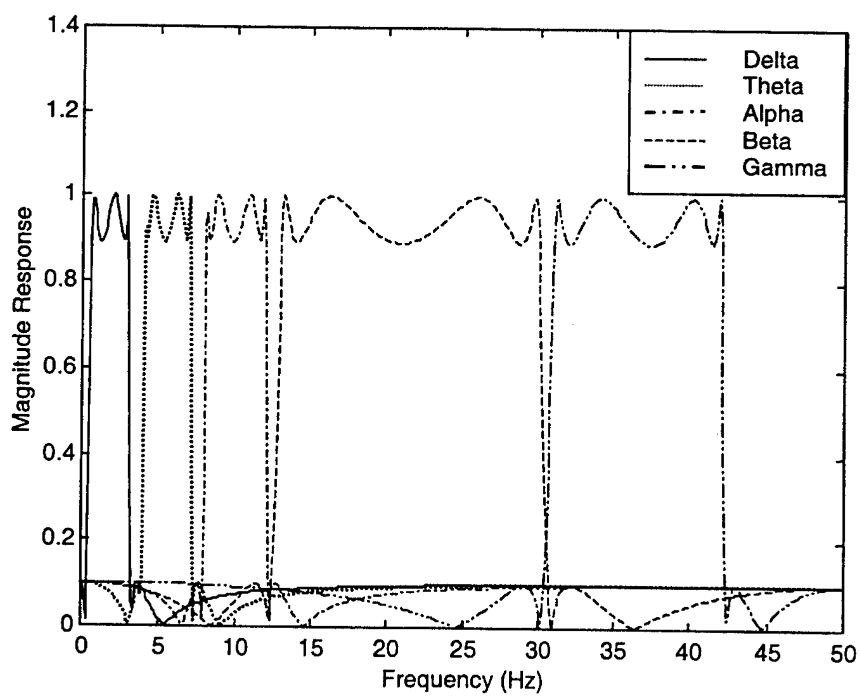


Figure 2. Frequency Response of Elliptical Filters

The data from the middle five minutes of each workload level were segmented into ten-second windows with a fifty-percent overlap as shown in Figure 3. Parseval's Theorem states that the integral of the magnitude square of a time series is equal to the integral of the magnitude square of that time series' Fourier coefficients. In other words, the energy in the time domain is equivalent to the energy in the frequency domain. Making use of this theorem, we determined the log power of each band using

$$P = 10 * \log\left(\sum f(t)^2\right) \tag{1}$$

Log power of delta, theta, alpha, beta and gamma bands from the 17 sites were used resulting in 85 features. Three peripheral physiological features were also used. The log power of the EOG channel and the average heart rate and average respiration rate completed the battery of 88 features.

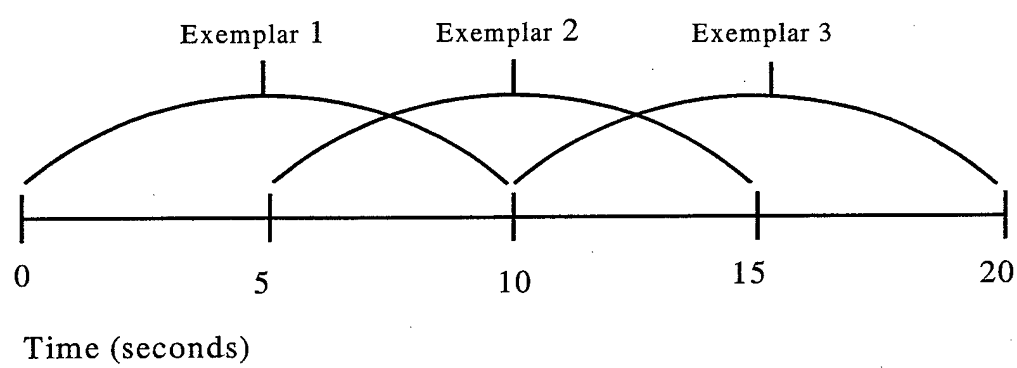


Figure 3. Description of Moving Window

Artificial Neural Network

A feedforward backpropagation ANN was used in this study (Widrow and Lehr, 1990; Lippmann, 1987). A backpropagation ANN classifier maps input vectors to output vectors in two phases. First, the network learns the input-output classification from a set of training vectors. Then, after training, the network acts as a classifier for new vectors.

The backpropagation algorithm initializes the network with a random set of weights for each fully connected layer, then the network trains using the input-output pairs. The learning algorithm uses a two-stage process for each pair: forward pass and backward pass. The forward pass propagates the input vector through the network until it reaches the output layer. First, the input vector propagates to the hidden units. Each hidden unit calculates the weighted sum of the

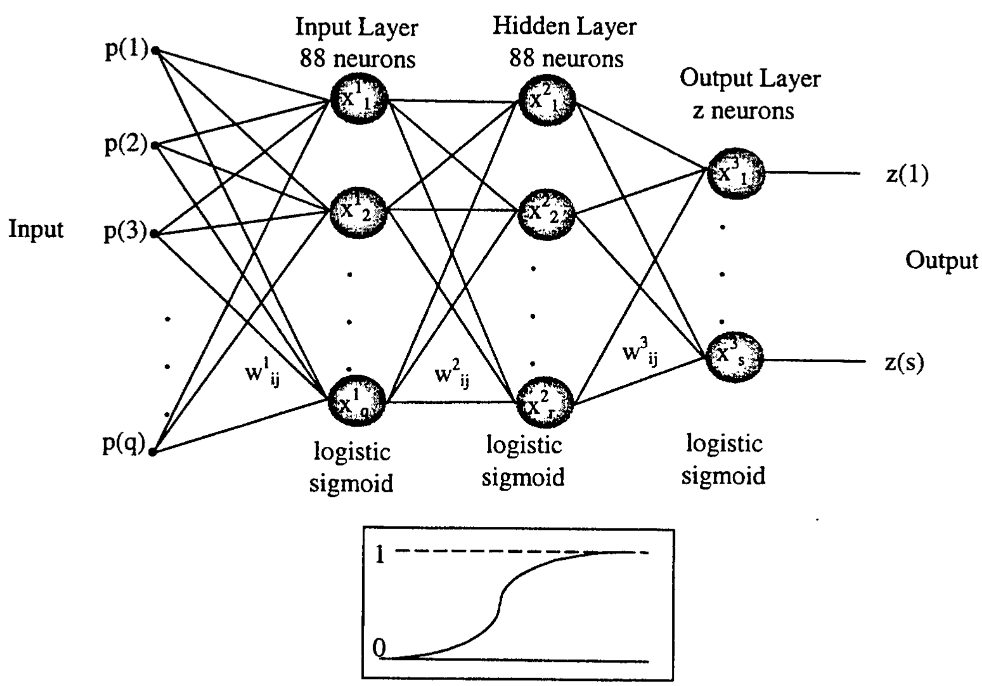


Figure 4. Network architecture showing a fully connected network with the number of neurons in each layer. The form of the logistic sigmoid activation function is provided at the bottom.

input vector and its associated interconnection weights. Each hidden unit uses the weighted sum to calculate its activation. Next, hidden unit activation propagates to the output layer. Each node in the output layer calculates its weighted sum and activation. Figure 4 shows the forward pass and Figure 5 is a typical unit featuring the summation and the activation. The output of the network is compared to the expected output of the input-output pairs; and their difference defines the output error. In the second stage of network training, the output error propagates backward to update the network weights. First, the error passes from the output layer to the hidden layer updating output weights. Next, each hidden unit calculates an error based on the error from each

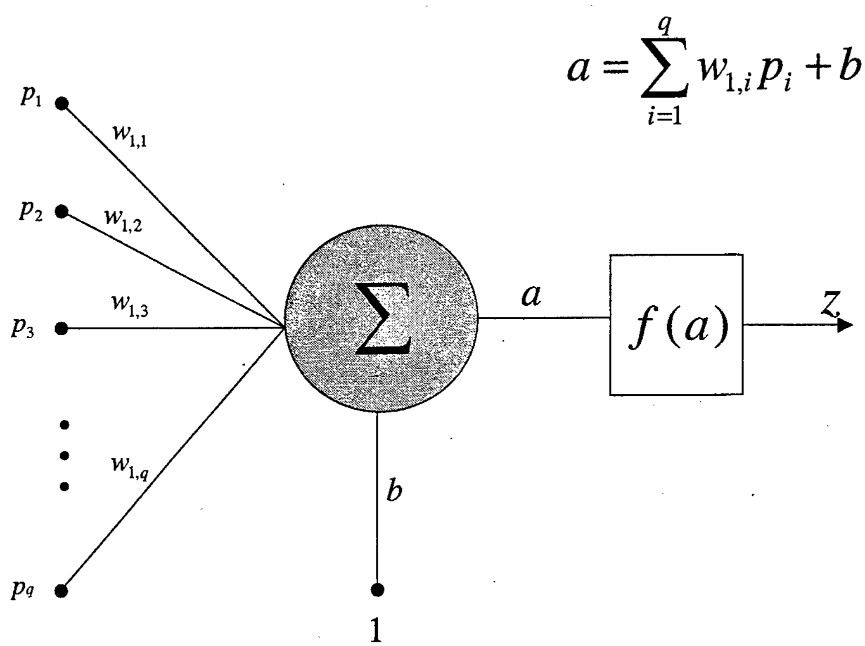


Figure 5. Individual neuron showing the weighted sum of the inputs followed by the logistic sigmoid activation function, *f(a)*.

output unit. The error from the hidden units updates the input weights. One training epoch passes when the network sees all the input-output pair in the training set. Training stops when the sum-squared error is acceptable or when a predefined number of epochs passes. The algorithm (backward pass) attempts to minimize the error or energy function

$$E = \sum_{i=1}^{m} \left\| \bar{z}_i - \bar{t}_i \right\|^2 , \tag{2}$$

where m is the size of the training set, z is the neural network output vector, and t is the expected output for each training input-output pair i.

It may be simpler to examine the algorithm as a series of steps. The steps for implementing a backpropagation neural network are as follows (Lippmann, 1987):

(1) Initialize the weights (w_i) and biases (b_i) where i is the current iteration.

(2) Present the input matrix (p) and the target vector (t).

(3) Calculate the output of the network (z_i).

(4) Calculate the error ($e = z_i - t$).

(5) Determine the new weights (w_{i+1}) where $i+1$ is the next iteration.

(6) Determine the new learning rate.

(7) Repeat steps 2 through 5 until desired error is achieved.

Mathematically, these steps were as follows: (Haykin, 1999; Widrow and Stearns, 1985; Widrow and Lehr, 1990). The weights and biases were initialized using a random number generator and limiting the values to the range –0.5 to 0.5, which is the nearly linear region of the hyperbolic sigmoid activation function.

The input data were normalized between 0 and 1 using a min-max equation

$$pn(i) = \frac{p(i) - p_{\min}}{p_{\max} - p_{\min}}, \tag{3}$$

where pn is the normalized input vector, p is the input vector, $p_{\min}$ and $p_{\max}$ are the minimum and maximum values for each feature, and i represents the i^{th} exemplar. The target vectors were assigned based on the *a priori* target output class. The class target output neuron was assigned 0.9 and all other target output neurons were assigned 0.1. The target vectors were [0.9 0.1 0.1 0.1] for low workload, [0.1 0.9 0.1 0.1] for medium, [0.1 0.1 0.9 0.1] for high and [0.1 0.1 0.1 0.9] for the overload condition.

The output of the ANN is determined by propagating the normalized input through each layer of the backpropagation neural network. It is necessary to examine the output of an individual neuron and then expand that understanding to the framework of the entire network. As shown in Figure 5, the output of the individual node or neuron is

$$\mathrm{z} = \mathrm{f}(\mathrm{a}), \tag{4}$$

and

$$a = \sum_{j=1}^{q} \left(w_{1j} p_j + b \right), \tag{5}$$

where w_{1j} is the weight, p_j is the input and b is the bias, and $f(a)$ is the activation function acting on a. The figure suggests this neuron is in the input layer since the leading index on the weight is 1. Generalizing to any neuron results in

$$\mathrm{z}_j = \mathrm{f}(\mathrm{a}_j) \tag{6}$$

and

$$a_j = \sum_{j=1}^{q} \left(w_{ij} p_j + b_j \right). \tag{7}$$

Activation functions can be linear or nonlinear. A common activation function is a sigmoidal nonlinearity. In our case, it is a logistic sigmoid function with an output range $0 \leq f(a) \leq 1$ in the form

$$f(a) = \frac{1}{1+e^{-a}}. \tag{8}$$

The error is simply the difference between the output of the network and the expected target value.

$$E_k = \sum_{i=1}^{s} (z_i - t_i)^2 \tag{9}$$

where k is the error for the current input exemplar.

We can adjust the weights and try to minimize the error E_k through the backward path. Although the activation function is nonlinear, it is differentiable and we can compute $\frac{\partial E_k}{\partial w_{ij}}$ which we will make use of in our selection of a learning rule. The network algorithm is an extension of the Widrow-Hoff learning rule (Widrow and Lehr, 1990) which is a gradient descent algorithm based on Widrow's earlier work in Adaline and Madaline neural networks. This rule adjusts the weights using a method of steepest descent algorithm.

$$w_{ij}(n) = w_{ij}(n-1) - \mu \frac{\partial E}{\partial w_{ij}} \tag{10}$$

where μ is a constant that controls the speed of convergence (learning rate).

Adaptive learning and momentum were used to decrease the time required for training the networks and to ensure the network reaches a global mimina. Typically, gradient descent

methods use a fixed learning rate to control the rate of convergence. However, it is difficult to determine an optimum rate. If the fixed learning rate is too large, the gradient descent algorithm becomes unstable due to oscillations. If the learning rate is too small, the incremental steps along the error surface are small and in turn the algorithm takes a long time to converge to the desired error. Adapting the learning rate to optimize the learning progress can maintain stability while keeping the learning rate as large as possible to improve the rate of convergence. As the slope of the local error surface increases, the learning rate decreases to control stability.

Momentum prevents the network algorithm from becoming trapped in a local minimum. Essentially the algorithm will "jump over" or ignore small perturbations in the error surface. Modification of the delta learning rule to include momentum results in a new learning rule

$$w_{ij}(n) = \alpha w_{ij}(n-1) - \mu \frac{\partial E}{\partial w_{ij}}, \tag{11}$$

where α is the momentum and μ is the learning rate.

This process is repeated until a desired error is achieved. The desired error is problem specific and must be determined. We determined our target or desired error by the validation method. The neural nets were optimized by a validation method. The data were segmented into three data sets: a training data set, a validation set and a test data set. During training, the neural network adjusted the weights and biases based on the training data set. After each adjustment the weights were tested on the validation set and once the network reached a minimum solution the test set was used to evaluate the final weights. The training and the validation error initially follow the same path until the ANN begins to learn the idiosyncrasies of the training data set. The error for the training data set still continues to decrease after this point but the validation data set error increases due to the neural network overlearning the peculiarities of the training

data. The ideal stopping point for training is the minimum validation error. The ANNs were trained to a sum-squared error of 0.04 which generally occurred 10,000 epochs or passes through the data. These training criteria were used for the remainder of the analysis.

Once trained, ANN weights are fixed and the net acts as a pattern classifier. As a classifier, the ANN examines input vectors it has never seen and predicts the class of the input vector.

The number of nodes in the input layer, the hidden layer and the output layer defines the ANN used in this study (see Figure 4). The number of input units and the number of output units are problem dependent. In our case, the input layer consists of 88 neurons representing the 88 features which form the input space. The output layer consisted of two, four or seven neurons since the number of classes existing in the data determined the size of the output layer. The number of hidden units required is usually not known. Hidden units are the key to network learning and force the network to develop its own internal representation of the input space. The ANN that produces the best classification with the fewest units is selected as the best topology. A net with too few hidden units cannot learn the mapping to the required accuracy since the smaller hidden layer would limit interaction of the input space. Too many hidden units allow the net to 'memorize' the training data and will not generalize well to new data. We used 88 neurons in the hidden layer.

Feature Reduction

An important consideration in classification is determining the input features. This is essential for any classification problem or algorithm, be it nonlinear (ANNs) or linear (stepwise

discriminant analysis). Some input features may be redundant because they are highly correlated or duplicated with only scalar differences. Others may not provide any useful information for discrimination (noise). Decreasing the number of input features by removing the redundant or meaningless inputs reduces the computation required for training. Reducing the number of features also reduces the number of exemplars or samples necessary for adequate learning by the classification algorithm. The number of samples required to estimate the free parameters of the network model increases nonlinearly as the number of inputs increases. This increase is the 'curse of dimensionality' inherent to all pattern recognition models. As the number of dimensions increase the number of training data necessary to develop an adequate model is boundless.

The Ruck saliency measure (Ruck, Rogers and Kabrisky, 1990) was used to determine which features provide information for the classification algorithm. This technique calculates the partial derivative of each layer and rank orders the features based on the saliency measure. In essence, this method provides an input-output relationship between the network output layer and the input features. This partial derivative method is possible because the activation function is nonlinear but is differentiable. The derivative of the activation function (equation 8) used in this study is

$$f'(a) = f(a)(1 - f(a)). \tag{12}$$

Feature saliency is based on the concept that a fully trained network contains all the information for describing the relative importance or saliency of each of the input features. The partial derivatives look cumbersome but can be readily calculated using the chain rule and are easily implemented in vector form. These calculations are performed starting with the output layer. The partial derivative for the output layer is

$$\gamma_{k3}^{3} = f'(a_{k3}^{3}) \tag{13}$$

$$= a_{k3}^{3}(1 - a_{k3}^{3}), \tag{14}$$

where k3 represents each output neuron and, in our case, the output layer is the third layer. Recall from equation 7, *a* represents the weighted sum of the inputs to the activation function plus the bias or threshold. The second or hidden layer is a little more complicated:

$$\gamma_{k2}^{2} = f'(a_{k2}^{2})\sum_{k2}\gamma_{k2}^{3}w_{k2}^{3} \tag{15}$$

$$= a_{k2}^{2}(1 - a_{k2}^{2})\sum_{k2}\gamma_{k2}^{3}w_{k2}^{3} \tag{16}$$

In this case, k2 represents the second layer neurons. The input layer has the same form as the second or hidden layer:

$$\gamma_{k1}^{1} = f'(a_{k1}^{1})\sum_{k1}\gamma_{k1}^{2}w_{k1}^{2} \tag{17}$$

$$= a_{k1}^{1}(1 - a_{k1}^{1})\sum_{k1}\gamma_{k1}^{2}w_{k1}^{2} \tag{18}$$

Finally the partial derivative for the entire neural network is

$$\frac{\partial z_j}{\partial x_i} = \sum_{k1}\gamma_{k1}^{1}w_i^{1}. \tag{19}$$

Combining equations 13 through 19.

$$\frac{\partial z_j}{\partial x_i} = \sum_{k1}\left[a_{k1}^{1}(1 - a_{k1}^{1})\sum_{k1}\left[a_{k2}^{2}(1 - a_{k2}^{2})\sum_{k2}\left[a_{k3}^{3}(1 - a_{k3}^{3})\right]w_{k2}^{3}\right]w_{k1}^{2}\right]w_i^{1} \tag{20}$$

Once the partial derivatives have been calculated the saliency can be determined for each feature by

$$\Gamma_i = \sum_{p}\sum_{j}\left|\frac{\partial z_j}{\partial x_i}\right|, \tag{21}$$

where Γ_i is the saliency for the ith feature, j ranges over the outputs, and p ranges over the exemplar vectors in the training set.

Feature reduction was accomplished with an iterative approach. Multiple networks were trained using all the features described in the feature selection portion of this paper. The partial derivative saliency was calculated for each feature. The features were then rank ordered based on the computed saliency. The least salient feature was removed from the input matrix and the networks were retrained using the reduced feature set. This sequence was repeated until the networks would no longer converge or the classification accuracy dropped well below the accuracy using all the features. The minimum data set is the smallest set that has the highest classification accuracy. Figure 6 shows the typical response for this iterative process. In this case 12 salient features are the minimum number required

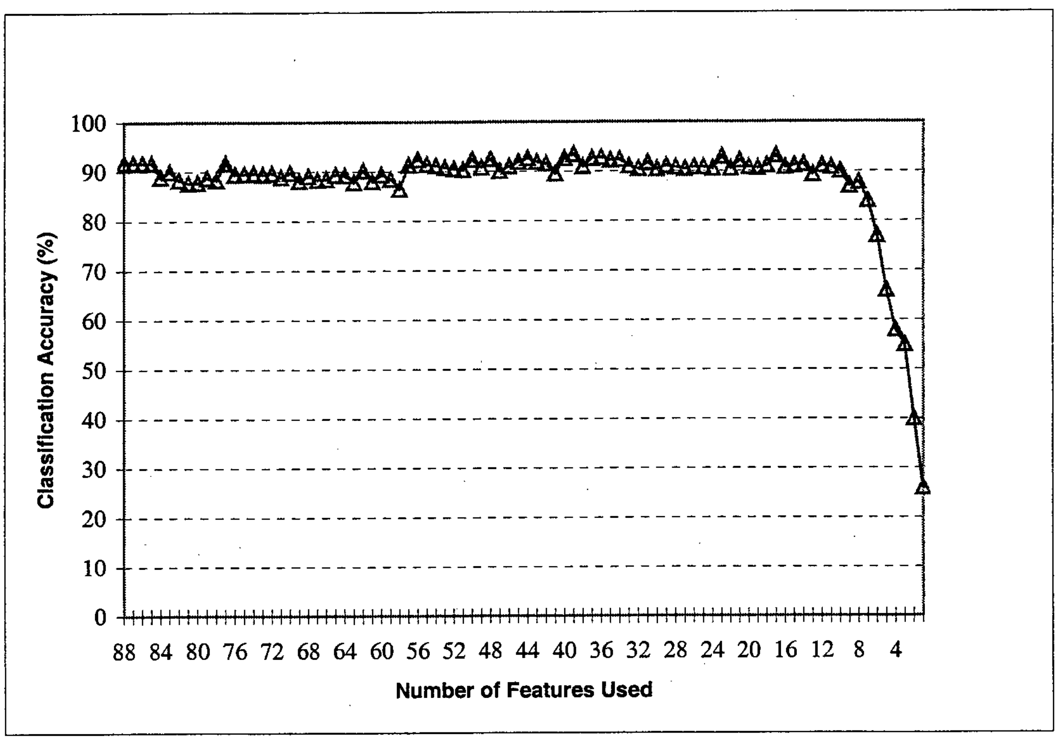


Figure 6. Classification behavior of the ANN as features are removed.

RESULTS

Seven-class Problem Using All Features

ANNs were trained with a seven-class problem using all available features. The volume and complexity data as well as the overload data for each subject were combined resulting in seven distinct classes. These classes were low complexity, medium complexity, high complexity, low volume, medium volume, high volume, and the overload condition. Table 1 shows the results of this comparison across subjects. The overload condition was correctly classified with the highest degree of accuracy at 90 percent correct. The classification accuracy across all conditions was 79.7%. The mean percentage correct for each subject is shown in Figure 7. The full individual subject results are located in Appendix C.

Although the workload states were labeled low, medium and high for both conditions, it is obvious that there were different classes between the volume and complexity conditions. This can be seen by how well the ANNs were able to differentiate between the conditions and workload levels. For example, the volume low is distinct from the complexity low. The volume low was correctly classified at 82 percent and was misclassified as complexity low only 4 percent. The remaining levels are similarly distinct.

Table 1: Seven-class Probability Matrix

	VL	VM	VH	CL	CM	CH	OL
VL	0.82	0.05	0.03	0.04	0.02	0.02	0.01
VM	0.05	0.73	0.11	0.02	0.03	0.04	0.02
VH	0.01	0.09	0.78	0.01	0.03	0.05	0.03
CL	0.06	0.02	0.01	0.82	0.07	0.03	0
CM	0.04	0.03	0.05	0.07	0.73	0.06	0.02
CH	0.03	0.04	0.05	0.02	0.05	0.78	0.03
OL	0.01	0.02	0.02	0	0.02	0.02	0.90

VL-Volume Low VM-Volume Medium VH- Volume High CL-Complexity Low CM-Complexity Medium CH-Complexity High OL-Overload

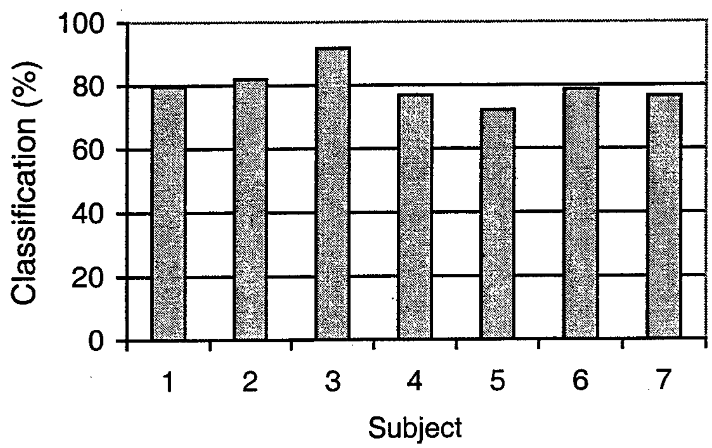


Figure 7. Individual Classification Accuracy

Four-class Problem Using All Features

When the volume and complexity data were analyzed separately, the classification accuracy across subjects and across difficulty levels using all features was 85.6 percent for the volume data and 83.4 percent for the complexity data. The condition with the highest rate of

accurate classification was the overload condition in both the volume and complexity conditions with 93.8 and 89.4 percent correct classification, respectively. All other conditions had classification accuracies around eighty percent. Tables 2 and 3 show the probability matrices for both the volume and the complexity data.

Table 2: Complexity Data Probability Matrix

	Test Low	Test Medium	Test High	Test Overload
Truth Low	0.8288	0.0886	0.0681	0.0145
Truth Medium	0.0804	0.7897	0.0847	0.0451
Truth High	0.0432	0.0811	0.8228	0.0529
Truth Overload	0.0185	0.0346	0.0525	0.8974

Table 3: Volume Data Probability Matrix

	Test Low	Test Medium	Test High	Test Overload
Truth Low	0.8545	0.0619	0.0655	0.0182
Truth Medium	0.0543	0.8096	0.1124	0.0237
Truth High	0.0419	0.1155	0.8213	0.0213
Truth Overload	0.0131	0.0287	0.0200	0.9383

The overall classification accuracy across subjects was nearly equivalent between the volume and the complexity conditions. The volume data indicates the ANNs were more likely to confuse the medium and high workload conditions. Medium workload was classified correctly 81% of the cases and was misclassified as high workload in 11% of the cases. The high workload condition was classified similarly with 82% classified correctly and 12% misclassified as medium workload. The complexity condition did not show the same behavior. The misclassification was distributed evenly between low and high workload for the medium workload condition while the low and high workload condition was misclassified more as medium workload. In all cases, correct classification was significantly above chance, which is 25% for the four conditions.

The average classification accuracy of the individual subjects ranged from 74.2 to 91.4 percent. Figure 8 shows the classification results for the individual subjects for both the volume and complexity data. Figure 9 shows the classification accuracy of the overload condition for each subject. In most subjects the classification accuracy for the overload condition was over 90 percent. The complete individual subject results can be seen in Appendix A.

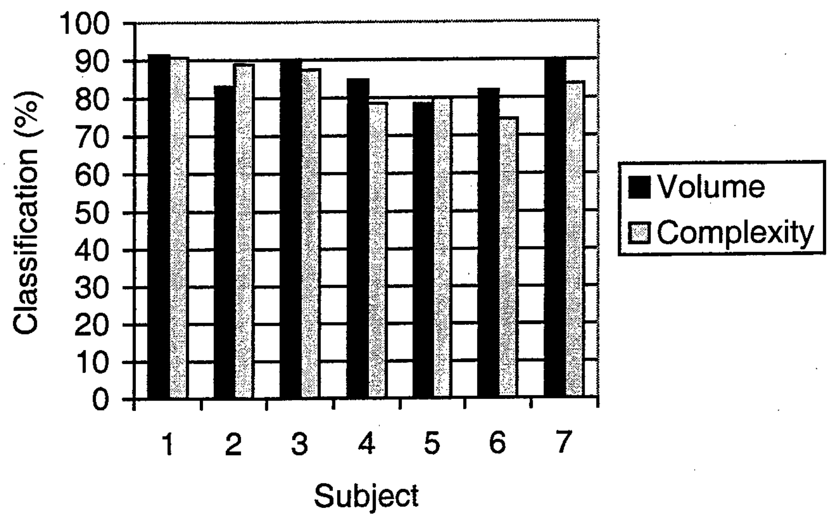


Figure 8. Individual Subject Classification Accuracy

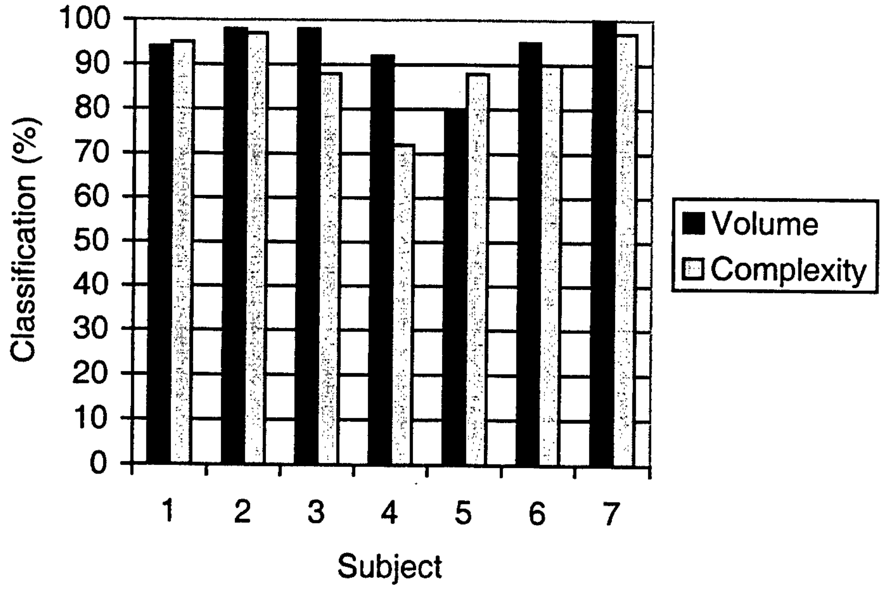


Figure 9. Overload Condition Classification Accuracy

Four-class Problem Using Reduced Features

The overall classification accuracies for both the volume and complexity data sets for each subject after feature reduction are shown in Figure 10. The across-subject average classification accuracy was very similar for both the volume and complexity data sets as shown in Tables 4 and 5. The results were 92.5% for the volume data and 92.6% for the complexity data. The individual subject results can be seen in Appendix B. Although the classification results were similar, the number of features required was different. The volume data required 22 features across subjects while the complexity data required 17 features across subjects to achieve the same results as shown in Tables 6 and 7. The ranking of features across condition was determined by the average saliency for each feature across condition. The results were then rank ordered and are listed in Table 8.

Table 4: Complexity Probability Matrix				
	Test Low	Test Medium	Test High	Test Overload
Truth Low	0.9258	0.0474	0.0193	0.0075
Truth Medium	0.0314	0.8904	0.0560	0.0222
Truth High	0.0147	0.0473	0.9147	0.0233
Truth Overload	0.0006	0.0107	0.0152	0.9735

Table 5: Volume Probability Matrix				
	Test Low	Test Medium	Test High	Test Overload
Truth Low	0.9052	0.0511	0.0320	0.0117
Truth Medium	0.0290	0.8964	0.0487	0.0259
Truth High	0.0205	0.0440	0.9249	0.0105
Truth Overload	0.0057	0.0145	0.0070	0.9728

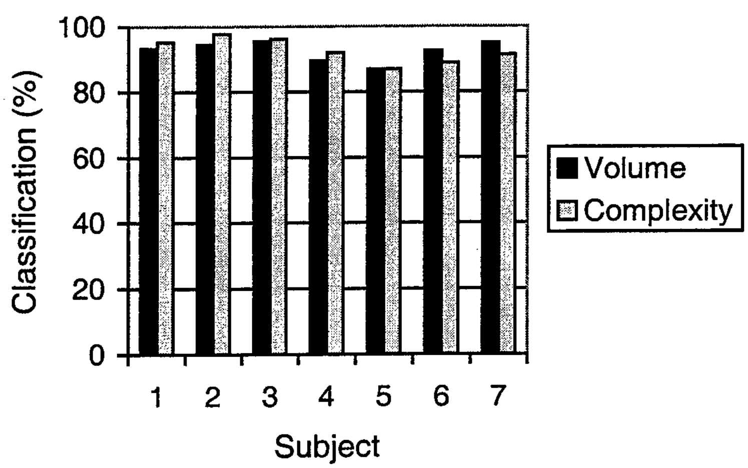


Figure 10. Subject classification accuracy after reduction for the four-class problem

Table 6: Salient Features – Volume

Feature	Mean Relative Saliency
T4 beta	0.044
T3 beta	0.043
O1 beta	0.037
C4 beta	0.035
F8 beta	0.034
T6 gamma	0.031
T6 beta	0.030
P4 beta	0.029
O2 beta	0.027
C4 delta	0.027
F4 gamma	0.026
F7 beta	0.025
Avg Hrtrate	0.024
C3 delta	0.024
T5 beta	0.023
FZ gamma	0.021
CZ delta	0.020
T5 theta	0.020
O1 alpha	0.020
PZ theta	0.018
F3 beta	0.017
PZ delta	0.016

Table 7 : Salient Features – Complexity

Feature	Mean Relative Saliency
O1 gamma	0.063
O2 gamma	0.057
O2 beta	0.049
O1 beta	0.047
T5 gamma	0.044
T5 beta	0.042
O1 alpha	0.039
T6 gamma	0.034
O2 alpha	0.031
T6 beta	0.030
PZ beta	0.027
C4 beta	0.024
Avg Hrtrate	0.022
C3 beta	0.021
T5 theta	0.021
F3 beta	0.021
F4 beta	0.020

Table 8 : Salient Features – Volume and Complexity

Feature	Mean Relative Saliency
O1 beta	0.042
O2 beta	0.038
O1 gamma	0.035
O2 gamma	0.035
T6 gamma	0.033
T5 beta	0.033
T4 beta	0.031
T6 beta	0.030
C4 beta	0.030
T5 gamma	0.030
O1 alpha	0.029
T3 beta	0.028
Avg Hrtrate	0.023
F8 beta	0.023
T5 theta	0.021
F3 beta	0.019
P4 beta	0.019
O2 alpha	0.018
F7 beta	0.017

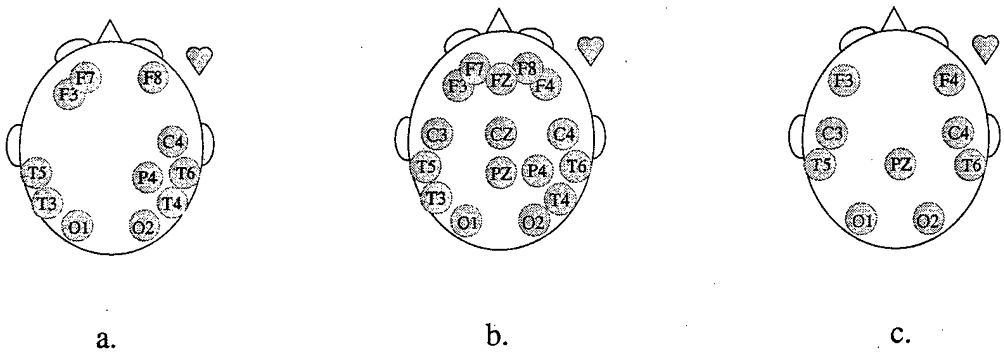


Figure 11. a) Salient electrode sites for four-class problem. b) Salient electrode sites for volume four-class problem. c) Salient electrode sites for complexity four-class problem

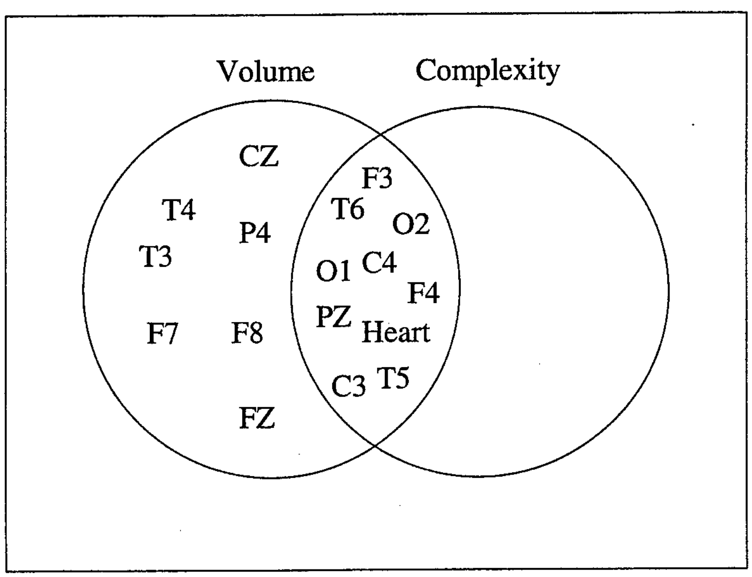


Figure 12. Overlap of electrode sites for four-class reduced feature analysis

Figure 11 shows the electrode placement of the salient features across conditions. The Venn diagram in Figure 12 indicates the electrode placement and the overlap between conditions. The volume condition ANNs used the same electrodes as the complexity condition plus seven additional electrode sites. The volume condition required more electrodes to have enough information to separate the workload classes.

Two-class Problem Using Reduced Features

The success of the artificial networks in distinguishing the overload condition at such a high level prompted the segmentation of the data into two groups; overload and not overload. The not overload condition is the balanced combination of the low, medium and high workload conditions. In other words, a third of the not overload condition came from each of the three workload conditions. The ANNs were trained with the two-class problem and the results are shown in Table 9 for the volume data and Table 10 for the complexity data. The individual subject average classification accuracy is shown in Figure 13. These results are after reducing the features using the partial derivative method of feature reduction.

Table 9: Two-class Volume Accuracy

	Test Not Overload	Test Overload
Truth Not Overload	0.9935	0.0065
Truth Overload	0.0367	0.9633

Table 10: Two-class Complexity Accuracy

	Test Not Overload	Test Overload
Truth Not Overload	0.9884	0.0116
Truth Overload	0.0369	0.9631

Both the volume and complexity conditions had high classification accuracies and used an average of eight salient features. The volume data set had an overall classification accuracy of 98.6% using an average of eight features. The complexity data had an overall classification accuracy of 98.2% using an average of eight features. These features are O1 beta, O2 beta, O2 gamma, O1 gamma, T6 beta, T5 beta, T4 beta, and T3 beta in descending order of importance as shown in Table 11. It is interesting to note that the important features are EEG and that no peripheral measures were used, as shown in Tables 12 and 13. Note the two most salient features were identical for both task condition manipulations. The two most salient features for both the volume and complexity conditions were O1 and O2 beta. Figure 11 illustrates the electrode placement of the most salient features. Also the higher frequency components of the EEG were used, namely the beta and gamma frequency bands.

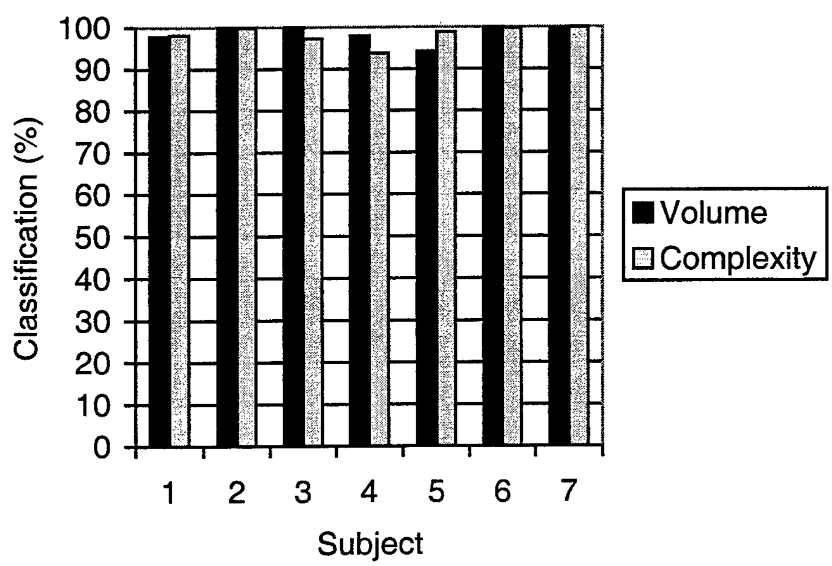


Figure 13. Individual subject classification accuracy for the two-class problem.

Table 11. Salient Features – Volume and Complexity	
Feature	Mean Relative Saliency
O1 beta	0.151
O2 beta	0.081
O2 gamma	0.052
O1 gamma	0.045
T6 beta	0.045
T5 beta	0.037
T4 beta	0.037
T3 beta	0.035

Table 12. Salient Features - Volume	
Feature	Mean Relative Saliency
O1 beta	0.198
O2 beta	0.070
T3 beta	0.048
PZ delta	0.047
PZ gamma	0.040
T4 beta	0.035
T5 beta	0.033
T3 theta	0.033

Table 13. Salient Features – Complexity	
Feature	Mean Relative Saliency
O1 beta	0.052
O2 beta	0.046
O2 gamma	0.045
O1 gamma	0.039
T6 beta	0.037
T6 theta	0.024
T5 beta	0.021
T4 beta	0.019

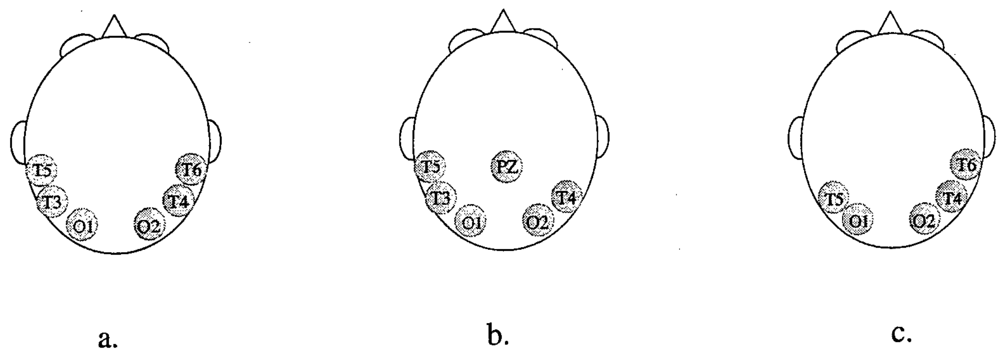


Figure 14. a) Salient electrode sites volume and complexity. b) Salient electrode site for volume data. c) Salient electrode site for complexity data.

The salient electrode placement for both the volume and complexity conditions was the same with one exception. The volume manipulation indicates the Pz electrode was used in addition to the others to determine the classes for this condition. This is shown in Figures 14 and 15. The salient EEG bands were all in the higher frequency bands, beta and gamma, except for one occurrence of the theta band. This pattern is very different from the four-class problem. The salient electrodes appear scattered over the entire scalp in the case of the four-class problem while, with the two-class problem the salient electrodes are located on the back and sides of the head.

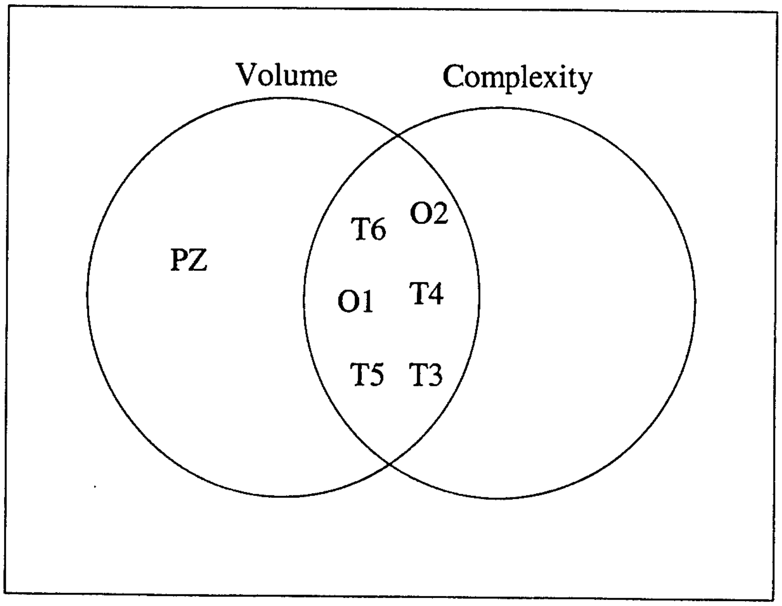


Figure 15. Overlap of Electrode Sites

DISCUSSION

The results of this study demonstrate that very high levels of classification accuracy can be achieved to discriminate among cognitive workload levels of air traffic controllers. Accuracies approaching 100 percent were found when only overload versus non-overload conditions were considered. The operator workload classification from the four level volume and complexity conditions also achieved very high levels of accuracy. Both manipulations of task difficulty yielded accuracies with an average correct classification of 92 percent. These procedures would be useful in both task evaluation and adaptive aiding situations. The results of the combined volume, complexity and overload conditions showed that the ANNs are very sensitive to the effects of performing different tasks.

Training and testing the artificial neural networks on all seven classes: low, medium and high for volume and complexity and the overload condition, showed that good classification accuracies could be achieved. The mean accuracy of 80 percent is much better than chance levels. The results also demonstrated that performing the volume and complexity manipulations produced separate operator states. There was very little overlap between the classification states between the two conditions. The incorrectly classified data were scattered among all of the other six conditions. That is, the low volume and complexity conditions were not misclassified any more frequently than with the other conditions. This indicates that the ANNs are very sensitive to the differences in the data. This high level of discrimination shows that the ANNs do not combine all similar levels of workload. The psychophysiological data contained sufficient differences between the various conditions that permitted the ANNs to successfully discriminate among them. On the other hand, this also means that the ANNs have to be trained on all

expected conditions since they did not generalize to other similar workload levels. It might be possible to combine the levels from the two conditions and achieve better results.

Higher levels of classification accuracy were achieved when the ANNs were trained on data from only the volume or complexity conditions. The mean accuracies increased to 86 and 83 percent correct for the volume and complexity conditions, respectively. While not extremely large, this increase in accuracy does show that the presence of less relevant data interferes with the accuracy of the ANNs. The increase in accuracy is a result of the 'curse of dimensionality' problem described above. The ANNs need more sample data to estimate the increased number of free variables (weights and biases) in the model. Feature reduction removes free variables from the model, therefore, allowing the algorithm to provide better estimates of the remaining variables using the fixed sample size.

Feature reduction using saliency analysis produced marked improvement in the classification accuracies. The mean percentage corrected jumped to 93 percent for both volume and complexity conditions. This seven to ten percent improvement is especially remarkable because this raised the correct classification levels to only seven percent, on the average, from perfect classification. These levels of classification accuracy reach the range where the procedures used here can be applied in actual work situations. The highest accuracy levels were found when the data were reduced to a two-class problem; overload versus combined low, medium and high. The overall classification accuracies were above 98 percent for both volume and complexity conditions. This near perfect accuracy shows the extremely high levels of discrimination that can be achieved with this type of data using ANNs and feature reduction. In task situations where cognitive load classification might be applied, this level of discrimination is necessary. That is, if the job difficulty and operator workload were manageable then intervention

by aiding options would not be required. However, if the workload is approaching or exceeding the capabilities of the operator then intervention would be necessary. If this near perfect classification accuracy can be replicated with different real world tasks and with larger numbers of operators then application to actual job situations will be considered.

Feature reduction using saliency analysis is an important element in the success of achieving such high levels of classification accuracy. The number of features eliminated increased from the four to the two-class problems. With the four-class problems 17 and 22 features of the original 88 were used. For the two-class problem the high accuracies were achieved with only eight of the original features. This reduction in feature size improved the ANNs by removing noise features and also improving the ANN node size to exemplar ratio which enhances classification accuracy. The smaller feature set reduces the amount of data required to completely estimate the parameters of the neural network. The electrode sites that contributed to the reduced feature set in the four-class problem were scattered over the scalp with a preponderance located around the edges of the scalp. The most salient electrode sites in the two-class problem were primarily located around the edges of the scalp. The EEG bands used in the reduced feature set were predominantly in the high EEG frequency bands, beta and gamma. It is possible that this could represent very low electromyographic activity from the muscles of the head.

These results are especially remarkable because professional air traffic controllers were engaged in a highly complex simulation. These are the conditions that are encountered in the real world. Typical laboratory studies use simple, often single task situations, and unselected subjects. While these procedures are required to provide rigid, highly controlled environments they are not typical of the usual day-to-day work environment. In the work environment, control

over the experimental conditions is often greatly reduced. Operators are engaged in several subtasks at the same time. The applicability of the laboratory results to the real world situation can be questioned. Further, subjects in most laboratory experiments are not selected for particular attributes nor are they highly trained on complex tasks. In many real world jobs operators are selected for attributes necessary for job performance and receive extensive training in order to become proficient at their jobs. In the study reported here these real world criteria were met. The air traffic controllers had been selected for training, successfully accomplished the required training and had several years experience on the job. This lends weight to the notion that these results can be successfully applied in complex real world jobs.

Before these procedures can be incorporated into day-to-day tasks the reliability of the procedures from day-to-day has to be demonstrated. The current results were based on data from only one day. These procedures must be robust enough to accommodate day-to-day variation before they can be widely accepted. Larger numbers of operators must be tested to determine if the same results can be achieved from everyone or if there are groups of operators on which these procedures do not work well enough. Additionally, faster ANN training procedures should be developed. It may be possible to develop a generic ANN that can be quickly adapted for each operator and save the lengthy training time when starting with an ANN with random initial weights. This generic ANN will have fixed weights that are 'closer' to the optimal solution than random weights. In other words, if we know the approximate final weights for good discrimination, then less time is required to determine the true weights since the algorithm is closer to the answer initially.

The overlap of the salient electrodes in both the two-class problem and the four-class problem indicates a small set of electrodes is essential in separating the classes. The electrodes

necessary in classifying workload levels for the volume condition for both the two and four-class cases contain the subset of electrodes used by the complexity manipulation. For example, in the two-class case, the six electrodes T3, T4, T5, T6, O1 and O2 are used for both the volume and the complexity manipulation of the air traffic control task. Electrode site has to be investigated to determine if universal sites can be used or if sites and EEG bands need to be selected for different job types and operators.

Saliency analysis procedures can also be used in research to help determine the mechanisms that underlie job performance and workload. The saliency analysis isolates those features in the data that contain the most variance. By examination of these data, insights can be gained into the relationships between job performance and brain and peripheral nervous system activity. This information could be used to improve workstations, work procedures and training procedures. The addition of performance features could improve the accuracy and reliability of ANN classification. For example, a measure of aircraft separation could indicate how well a controller is managing the airspace. Including performance and situational data could lead to better classification of cognitive load and an improved understanding of the underlying mechanisms of cognitive load.

Development of non-invasive or unobtrusive sensor technology and the use of neural networks will make real-time classification of cognitive workload in any environment a reality. The advent of high speed computer processors and the reduction in size of computer hardware will make possible the development of small, wearable operator functional state devices.

REFERENCES

Akay, M. (1994). *Biomedical Signal Processing*, Academic Press, San Diego, CA.

Anderson, C.W., Devulapalli, S. and Stolz, E.A. (1995). Determining mental state from EEG signals using parallel implementations of neural networks, *Scientific Programming, Special Issue on Applications Analysis*, Vol. 4, No. 3, pp. 171-183.

Benel, R.A., Dancey, R.D., Dehn, J.D., Gutmann, J.C. and Smith, D.M. (1989). Advanced automated systems design, *Proceedings of the IEEE*, Vol. 77, No. 11, pp. 1653-1660.

Brookings, J. B., Wilson, G. F., and Swain, C. R. (1996). Physiological responses to changes in workload during simulated air traffic control, *Biological Psychology*, Vol. 42, pp. 361-377.

Demuth, H. and Beale, M. (1998). *Neural Network Toolbox User's Guide* Version 3, MathWorks, Natick, MA.

Gevins, A., Smith, M.E., Leong, H., McEvoy, L., Whitfield, S., Du, R., and Rush, G., (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition methods, *Human Factors*, vol. 40, no. 1, pp. 79-91.

Greene, K.A., Bauer, K.W, Kabrisky, M., Rogers, S.K., Russell, C.A., and Wilson, G.F. (1996). A preliminary investigation of selection of EEG and psychophysiological features for classifying pilot workload, In Dagli, et al. (Eds.), *Smart Engineering Systems: Neural Networks, Fuzzy Logic and Evolutionary Programming,* ASME Press, pp. 691-697.

Hart, S.G. and Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In P.A. Hancock and N. Meshkati (Eds.), *Human Mental Workload*, Amsterdam: North-Holland.

Haykin, S., (1999). *Neural Networks: A Comprehensive Foundation*, 2nd Ed., Prentice Hall, Upper Saddle River, NJ.

Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, vol. 10, pp. 371-375.

Kramer, A.F. (1993). Physiological metrics of mental workload: a review of recent progress. In P. Ullsperger (Ed.) *Psychophysiology of Mental Workload*, Schriftenreibe der Bundesanstalt für Arbeftsmedizin, Sonderschriftz, Berlin, pp. 3-34.

Lee, K.L., Pawlak, W.S., Sanford, B.D. and Slattery, R.A., (1995). Improved navigational technology and air traffic control: a description of controller coordination and workload, *Proceedings of the Eighth International Symposium on Aviation Psychology*, vol. 1, pp. 444-449.

Lippmann, R.P. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine*, April 1987, pp. 4-22.

Morrison, R. and Wright, R.H., (1989). ATC control and communications problems: an overview of recent ASRS data, *Proceedings of the Fifth International Symposium on Aviation Psychology*, pp. 902-907.

Perry, T.S. (1997). In search of the future of air traffic control, *IEEE Spectrum*, Vol. 34, No. 8 pp. 18-35.

Reid, G.B. and Nygren, T.E. (1988). The subjective workload assessment technique: a scaling procedure for measuring mental workload. In P.A. Hancock and N. Meshkati (Eds.) *Human Mental Workload,* Amsterdam, The Netherlands: North Holland.

Ruck, D., Rogers, S. and Kabrisky, M., (1990). Feature selection using a multilayer perceptron, *Journal of Neural Network Computing*, vol. 2, no 2, pp. 40-48.

Russell, C.A., Monett, C.T. and Wilson, G.F. (1996). Mental workload classification using a backpropagation neural network. In Dagli, et al. (Eds.) *Smart Engineering Systems: Neural Networks, Fuzzy Logic and Evolutionary Programming,* ASME Press, 685-690.

Russell, C.A., Reid, G.B., and Vidulich, M.A., (2000). Pilot workload classification using artificial neural networks in a simulated SCUD hunt mission. *Proceedings of the 44th Annual Meeting of the Human Factors and Ergonomics Society*, 3, 109.

Widrow, B. and Lehr, M. (1990). 30 years of adaptive neural networks: perceptron, madaline and backpropagation, *Proceedings of IEEE*, Vol. 78, no. 9, pp. 1415-1442.

Widrow, B. and Stearns, S.D. (1985). *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ.

APPENDIX A – INDIVIDUAL SUBJECT PROBABILITY MATRICES USING ALL FEATURES *

Subject 1 – Complexity

	Low	Medium	High	Overload
Low	204	25	0	0
	0.89	0.11	0	0
Medium	19	211	2	7
	0.08	0.88	0.01	0.03
High	2	10	213	10
	0.01	0.04	0.91	0.04
Overload	3	1	8	225
	0.01	0	0.03	0.95

Subject 1 - Volume

	Low	Medium	High	Overload
Low	227	6	2	9
	0.93	0.02	0.01	0.04
Medium	6	202	24	1
	0.03	0.87	0.10	0
High	0	13	209	6
	0	0.06	0.92	0.03
Overload	1	1	10	203
	0	0	0.05	0.94

Subject 2 – Complexity

	Low	Medium	High	Overload
Low	197	24	17	2
	0.82	0.10	0.07	0.01
Medium	12	194	11	6
	0.05	0.87	0.05	0.03
High	8	12	216	4
	0.03	0.05	0.90	0.02
Overload	3	3	2	229
	0.01	0.01	0.01	0.97

Subject 2 - Volume

	Low	Medium	High	Overload
Low	210	24	9	3
	0.85	0.10	0.04	0.01
Medium	15	165	45	3
	0.07	0.72	0.20	0.01
High	2	49	193	1
	0.01	0.20	0.79	0
Overload	1	4	0	196
	0	0.02	0	0.98

Subject 3 - Complexity

	Low	Medium	High	Overload
Low	241	5	15	1
	0.92	0.02	0.06	0
Medium	0	197	9	2
	0	0.95	0.04	0.01
High	10	21	191	28
	0.04	0.08	0.76	0.11
Overload	0	2	25	193
	0	0.01	0.11	0.88

Subject 3 - Volume

	Low	Medium	High	Overload
Low	203	7	18	0
	0.89	0.03	0.08	0
Medium	1	198	29	0
	0	0.87	0.13	0
High	12	19	203	3
	0.05	0.08	0.86	0.01
Overload	0	2	2	243
	0	0.01	0.01	0.98

Subject 4 - Complexity

	Low	Medium	High	Overload
Low	189	10	18	6
	0.85	0.04	0.08	0.03
Medium	1	183	14	40
	0	0.77	0.06	0.17
High	14	2	189	29
	0.06	0.01	0.81	0.12
Overload	4	30	34	177
	0.02	0.12	0.14	0.72

Subject 4 - Volume

	Low	Medium	High	Overload
Low	194	22	20	4
	0.81	0.09	0.08	0.02
Medium	16	174	14	8
	0.08	0.82	0.07	0.04
High	11	19	190	8
	0.05	0.08	0.83	0.04
Overload	8	4	7	221
	0.03	0.02	0.03	0.92

Subject 5 - Complexity	Low	Medium	High	Overload
Low	187	27	9	10
	0.80	0.12	0.04	0.04
Medium	26	162	22	14
	0.12	0.72	0.10	0.06
High	2	40	192	8
	0.01	0.17	0.79	0.03
Overload	3	16	7	195
	0.01	0.07	0.03	0.88

Subject 5 - Volume	Low	Medium	High	Overload
Low	178	11	26	11
	0.79	0.05	0.12	0.05
Medium	13	175	19	26
	0.06	0.75	0.08	0.11
High	19	11	180	16
	0.08	0.05	0.80	0.07
Overload	7	31	9	188
	0.03	0.13	0.04	0.80

Subject 6 - Complexity	Low	Medium	High	Overload
Low	170	35	19	5
	0.74	0.15	0.08	0.02
Medium	43	142	60	4
	0.17	0.57	0.24	0.02
High	16	28	180	9
	0.07	0.12	0.77	0.04
Overload	11	3	9	206
	0.05	0.01	0.04	0.90

Subject 6 - Volume	Low	Medium	High	Overload
Low	209	7	26	2
	0.86	0.03	0.11	0.01
Medium	4	185	33	0
	0.02	0.83	0.15	0
High	21	65	166	0
	0.08	0.26	0.66	0
Overload	4	4	4	210
	0.02	0.02	0.02	0.95

Subject 7 - Complexity	Low	Medium	High	Overload
Low	187	21	35	0
	0.77	0.09	0.14	0
Medium	29	188	19	0
	0.12	0.80	0.08	0
High	20	22	189	0
	0.09	0.10	0.82	0
Overload	6	1	0	223
	0.03	0	0	0.97

Subject 7 - Volume	Low	Medium	High	Overload
Low	188	25	7	1
	0.85	0.11	0.03	0
Medium	32	198	16	0
	0.13	0.80	0.07	0
High	4	14	210	1
	0.02	0.06	0.92	0
Overload	0	0	0	244
	0	0	0	1

* Each table consists of each row signifying truth and each column representing testing. For each true level, the top number is the number of samples classified at each test level and the bottom number is the relative frequency of that occurrence.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B – INDIVIDUAL SUBJECT PROBABILITY MATRICES USING REDUCED FEATURES *

Subject 1 - Complexity

	Low	Medium	High	Overload
Low	212	9	2	0
	0.95	0.04	0.01	0
Medium	7	209	1	5
	0.03	0.94	0	0.02
High	0	6	229	5
	0	0.03	0.95	0.02
Overload	0	1	8	226
	0	0	0.03	0.96

Subject 1 - Volume

	Low	Medium	High	Overload
Low	213	13	0	5
	0.92	0.06	0	0.02
Medium	9	208	11	1
	0.04	0.91	0.05	0
High	2	17	216	3
	0.01	0.07	0.91	0.01
Overload	1	0	1	220
	0	0	0	0.99

Subject 2 - Complexity

	Low	Medium	High	Overload
Low	222	7	3	0
	0.96	0.03	0.01	0
Medium	2	224	5	3
	0.01	0.96	0.02	0.01
High	0	0	229	0
	0	0	1	0
Overload	0	0	0	225
	0	0	0	1

Subject 2 - Volume

	Low	Medium	High	Overload
Low	222	11	1	0
	0.95	0.05	0	0
Medium	11	199	8	12
	0.05	0.87	0.03	0.05
High	1	5	218	0
	0	0.02	0.97	0
Overload	0	2	0	230
	0	0.01	0	0.99

Subject 3 - Complexity

	Low	Medium	High	Overload
Low	227	2	1	0
	0.99	0.01	0	0
Medium	4	203	12	0
	0.02	0.93	0.05	0
High	1	3	219	7
	0	0.01	0.95	0.03
Overload	0	3	3	235
	0	0.01	0.01	0.98

Subject 3 - Volume

	Low	Medium	High	Overload
Low	228	2	5	0
	0.97	0.01	0.02	0
Medium	4	216	15	0
	0.02	0.92	0.06	0
High	4	11	223	0
	0.02	0.05	0.94	0
Overload	0	0	0	212
	0	0	0	1

Subject 4 - Complexity

	Low	Medium	High	Overload
Low	210	9	7	6
	0.91	0.04	0.03	0.03
Medium	2	230	1	7
	0.01	0.96	0	0.03
High	11	5	191	13
	0.05	0.02	0.87	0.06
Overload	0	3	10	215
	0	0.01	0.04	0.94

Subject 4 - Volume

	Low	Medium	High	Overload
Low	194	9	11	5
	0.89	0.04	0.05	0.02
Medium	3	203	14	17
	0.01	0.86	0.06	0.07
High	11	8	202	9
	0.05	0.03	0.88	0.04
Overload	0	2	7	225
	0	0.01	0.03	0.96

Subject 5 - Complexity	Low	Medium	High	Overload
Low	202	15	7	3
	0.89	0.07	0.03	0.01
Medium	11	179	15	18
	0.05	0.80	0.07	0.08
High	3	31	209	4
	0.01	0.13	0.85	0.02
Overload	1	10	3	209
	0	0.04	0.01	0.94

Subject 5 - Volume	Low	Medium	High	Overload
Low	189	29	20	8
	0.77	0.12	0.08	0.03
Medium	9	195	6	12
	0.04	0.88	0.03	0.05
High	6	7	208	5
	0.03	0.03	0.92	0.02
Overload	2	15	2	207
	0.01	0.07	0.01	0.92

Subject 6 - Complexity	Low	Medium	High	Overload
Low	198	25	4	3
	0.86	0.11	0.02	0.01
Medium	19	186	34	2
	0.08	0.77	0.14	0.01
High	4	3	211	9
	0.02	0.01	0.93	0.04
Overload	0	0	0	222
	0	0	0	1

Subject 6 - Volume	Low	Medium	High	Overload
Low	209	3	13	1
	0.92	0.01	0.06	0
Medium	2	210	12	0
	0.01	0.94	0.05	0
High	9	21	202	0
	0.04	0.09	0.87	0
Overload	6	0	1	231
	0.03	0	0	0.97

Subject 7 - Complexity	Low	Medium	High	Overload
Low	213	9	7	0
	0.93	0.04	0.03	0
Medium	6	215	23	1
	0.02	0.88	0.09	0
High	5	29	202	0
	0.02	0.12	0.86	0
Overload	0	0	0	210
	0	0	0	1

Subject 7 - Volume	Low	Medium	High	Overload
Low	216	16	2	0
	0.92	0.07	0.01	0
Medium	9	223	13	0
	0.04	0.91	0.05	0
High	0	2	222	0
	0	0.01	0.99	0
Overload	0	4	0	213
	0	0.02	0	0.98

* Each table consists of each row signifying truth and each column representing testing. For each true level, the top number is the number of samples classified at each test level and the bottom number is the relative frequency of that occurrence.

APPENDIX C – SEVEN-CLASS INDIVIDUAL SUBJECT PROBABILITY MATRICES *

Subject 1

	VL	VM	VH	CL	CM	CH	OL
VL	209	5	1	5	8	0	2
	0.91	0.02	0	0.02	0.03	0	0.01
VM	1	156	6	6	11	37	0
	0	0.72	0.03	0.03	0.05	0.17	0
VH	0	7	154	4	14	8	11
	0	0.04	0.78	0.02	0.07	0.04	0.06
CL	9	0	0	211	18	0	0
	0.04	0	0	0.89	0.08	0	0
CM	35	0	21	14	182	0	4
	0.14	0	0.08	0.05	0.71	0	0.02
CH	2	49	16	0	3	159	10
	0.01	0.21	0.07	0	0.01	0.67	0.04
OL	1	0	12	1	2	6	210
	0	0	0.05	0	0.01	0.03	0.91

VL-Volume Low VM-Volume Medium VH- Volume High CL-Complexity Low CM-Complexity Medium CH-Complexity High OL-Overload

Subject 2

	VL	VM	VH	CL	CM	CH	OL
VL	206	14	1	4	0	13	0
	0.87	0.06	0	0.02	0	0.05	0
VM	12	136	57	8	1	3	5
	0.05	0.61	0.26	0.04	0	0.01	0.02
VH	0	52	156	7	15	2	2
	0	0.22	0.67	0.03	0.06	0.01	0.01
CL	1	5	7	195	11	6	0
	0	0.02	0.03	0.87	0.05	0.03	0
CM	2	6	8	24	178	11	3
	0.01	0.03	0.03	0.10	0.77	0.05	0.01
CH	2	0	0	2	5	222	0
	0.01	0	0	0.01	0.02	0.96	0
OL	0	0	0	0	0	0	228
	0	0	0	0	0	0	1

VL-Volume Low VM-Volume Medium VH- Volume High CL-Complexity Low CM-Complexity Medium CH-Complexity High OL-Overload

Subject 3	VL	VM	VH	CL	CM	CH	OL
VL	213	5	8	6	0	0	0
	0.92	0.02	0.03	0.03	0	0	0
VM	5	190	34	1	0	0	0
	0.02	0.83	0.15	0	0	0	0
VH	2	17	199	2	1	0	6
	0.01	0.07	0.88	0.01	0	0	0.03
CL	0	0	1	227	3	3	0
	0	0	0	0.97	0.01	0.01	0
CM	0	0	0	4	215	8	0
	0	0	0	0.02	0.95	0.04	0
CH	0	0	1	0	10	220	5
	0	0	0	0	0.04	0.93	0.02
OL	0	0	0	0	8	2	214
	0	0	0	0	0.04	0.01	0.96

VL-Volume Low VM-Volume Medium VH- Volume High CL-Complexity Low CM-Complexity Medium CH-Complexity High OL-Overload

Subject 4	VL	VM	VH	CL	CM	CH	OL
VL	183	14	12	12	0	8	0
	0.80	0.06	0.05	0.05	0	0.03	0
VM	11	179	7	10	3	9	4
	0.05	0.80	0.03	0.04	0.01	0.04	0.02
VH	8	10	149	0	0	50	12
	0.03	0.04	0.65	0	0	0.22	0.05
CL	32	9	2	182	5	9	0
	0.13	0.04	0.01	0.76	0.02	0.04	0
CM	3	13	1	1	188	2	14
	0.01	0.06	0	0	0.85	0.01	0.06
CH	10	12	37	4	0	158	15
	0.04	0.05	0.16	0.02	0	0.67	0.06
OL	1	3	2	0	14	14	198
	0	0.01	0.01	0	0.06	0.06	0.85

VL-Volume Low VM-Volume Medium VH- Volume High CL-Complexity Low CM-Complexity Medium CH-Complexity High OL-Overload

Subject 5	VL	VM	VH	CL	CM	CH	OL
VL	181	25	11	12	5	2	12
	0.73	0.10	0.04	0.05	0.02	0.01	0.05
VM	10	150	5	9	15	4	27
	0.05	0.68	0.02	0.04	0.07	0.02	0.12
VH	4	5	191	1	9	9	3
	0.02	0.02	0.86	0	0.04	0.04	0.01
CL	14	12	1	177	26	1	0
	0.06	0.05	0	0.77	0.11	0	0
CM	7	16	19	24	138	11	7
	0.03	0.07	0.09	0.11	0.62	0.05	0.03
CH	6	3	30	3	32	164	6
	0.02	0.01	0.12	0.01	0.13	0.67	0.02
OL	13	30	11	0	13	8	148
	0.06	0.13	0.05	0	0.06	0.04	0.66

VL-Volume Low VM-Volume Medium VH- Volume High CL-Complexity Low CM-Complexity Medium CH-Complexity High OL-Overload

Subject 6	VL	VM	VH	CL	CM	CH	OL
VL	179	10	16	5	0	0	0
	0.85	0.05	0.08	0.02	0	0	0
VM	24	157	48	0	0	0	0
	0.10	0.69	0.21	0	0	0	0
VH	7	40	179	0	0	0	0
	0.03	0.18	0.79	0	0	0	0
CL	0	1	0	190	25	15	3
	0	0	0	0.81	0.11	0.06	0.01
CM	2	0	0	43	138	44	0
	0.01	0	0	0.19	0.61	0.19	0
CH	0	0	1	5	29	201	13
	0	0	0	0.02	0.12	0.81	0.05
OL	0	0	0	4	1	7	223
	0	0	0	0.02	0	0.03	0.95

VL-Volume Low VM-Volume Medium VH- Volume High CL-Complexity Low CM-Complexity Medium CH-Complexity High OL-Overload

	Subject 7						
	VL	VM	VH	CL	CM	CH	OL
VL	158	8	6	19	18	15	0
	0.71	0.04	0.03	0.08	0.08	0.07	0
VM	12	183	19	2	12	3	0
	0.05	0.79	0.08	0.01	0.05	0.01	0
VH	0	5	192	4	10	4	11
	0	0.02	0.85	0.02	0.04	0.02	0.05
CL	34	2	3	144	19	20	0
	0.15	0.01	0.01	0.65	0.09	0.09	0
CM	21	18	29	12	146	16	2
	0.09	0.07	0.12	0.05	0.60	0.07	0.01
CH	23	6	0	12	11	180	0
	0.10	0.03	0	0.05	0.05	0.78	0
OL	0	0	0	0	0	0	231
	0	0	0	0	0	0	1

VL-Volume Low VM-Volume Medium VH- Volume High CL-Complexity Low CM-Complexity Medium CH-Complexity High OL-Overload

* Each table consists of each row signifying truth and each column representing testing. For each true level, the top number is the number of samples classified at each test level and the bottom number is the relative frequency of that occurrence.